

ABSTRACT OF THE DISCLOSURE

The present invention relates to analysis of large, disk resident data sets using a Patient Rule Induction Method (PRIM) in a computer system wherein a relational data table is initially received. The relational data table includes continuous attributes, discrete attributes, a matter parameter and a cost attribute. The cost attribute represents cost output values based on continuous attribute values and discrete attribute values as inputs. A hyper-rectangle is then formed which encloses a multi-dimensional space defined by the continuous attribute values and the discrete attribute values. The continuous attribute values and the discrete attribute values are represented as points within the multi-dimensional space. A plurality of points along edges of the hyper-rectangle are then removed based on an average of the cost output value from the plurality of points until a count of the points enclosed within the hyper-rectangle equals the meta parameter. Discrete attribute values and continuous attribute values which were removed from the hyper-rectangle are next added along edges of the hyper-rectangle until a sum of the cost output value over the multi-dimensional space enclosed by the hyper-rectangle changes. In a further embodiment a parallel architecture computer system calculates the cost attribute average values over the plurality of points enclosed by the hyper-rectangle in parallel. The invention analyzes large disk resident data sets without having to load the data set into main memory and can be practiced on a parallel computer architecture or a symmetric multi-processor architecture to improve performance.